Chapter 2

Imaging and Image Representation

Humans derive a great deal of information about the world through their visual sense. Light reflects off objects and sometimes passes through objects to create an image on the retina of each eye. From this pair of images much of the structure of the 3D environment is derived. The important components are thus (a) a scene of objects, (b) illumination of the objects, and (c) sensing the illumination reflecting off the objects (or passing through them).

The major purpose of this chapter is to describe how sensors produce digital images of 2D or 3D scenes. Different kinds of radiation that reflect from or penetrate objects in the physical world can be sensed by different imaging devices. The 2D digital image is an array of intensity samples reflected from or transmitted through objects: this image is processed by a machine or computer program in order to make decisions about the scene. Often, a 2D image represents a projection of a 3D scene; this is the most common representation used in machine vision and in this book. At the end of the chapter, we discuss some relationships between structures in the 3D world and structures in the 2D image.

Various sections of this chapter are marked by an '*' to indicate that they provide technical details that can be skipped by a reader who is not particularly interested in them at this point.

2.1 Sensing Light

Much of the history of science can be told in terms of the progress of devices created to sense and produce different types of electro-magnetic radiation, such as radio waves, X-rays, microwaves, etc. The chemicals in the receptors of the human eye are sensitive to radiation (light) with wavelengths ranging from roughly 400 nanometers (violet) to 800 nanometers (red). Snakes and CCD sensors (see below) can sense wavelengths longer than 800 nanometers (infrared). There are devices to detect very short length X-rays and those which detect long radio waves. Different wavelengths of radiation have different properties; for example, X-rays can penetrate human bone while longer wavelength infrared might not penetrate

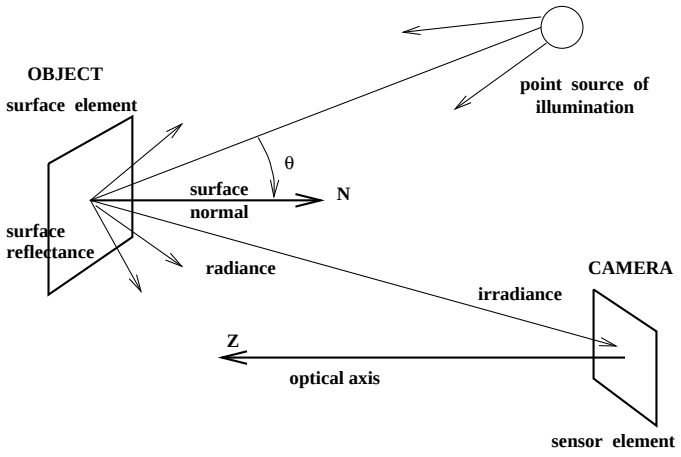

Figure 2.1: Reflection of radiation received from a single source of illumination.

even clouds.

Figure 2.1 shows a simple model of common photography: a surface element, illuminated by a single source (the sun or a flash bulb) reflects radiation toward the camera, which senses it via chemicals on film. More details of this situation are covered in Chapter 6. Wavelengths in the light range result from generating or reflecting mechanisms very near the surface of objects. We are concerned with many properties of electro-magnetic radiation in this book; however, we will usually give a qualitative description of phenomena and leave the quantitative details to books in physics or optics. Application engineering requires some knowledge of the material being sensed and the radiation and sensor used.

2.2 Imaging Devices

There are many different devices that produce digital images. They differ in the phenomena sensed as well as in their electro-mechanical design. Several different sensors are described in this chapter; the most common ones are discussed in this section, others are left to an optional section later in the chapter. Our intent is to disclose the important functional and conceptual aspects of each sensor, leaving most technical information to outside reading.

CCD cameras

Figure 2.2 shows a camera built using charge-coupled device (CCD) technology, the most flexible and common input device for machine vision systems. The CCD camera is very much like a 35 mm film camera commonly used for family photos, except on the image plane, instead of chemical film reacting to light, tiny solid state cells convert light energy into electrical charge. Each cell converts the light energy it receives into an electrical charge. All cells are first cleared to 0, and then they begin to integrate their response to the light energy falling on them. A shutter may or may not be needed to control the sensing time. The image plane acts as a digital memory that can be read row by row by a computer input process. The figure shows a simple monochrome camera.

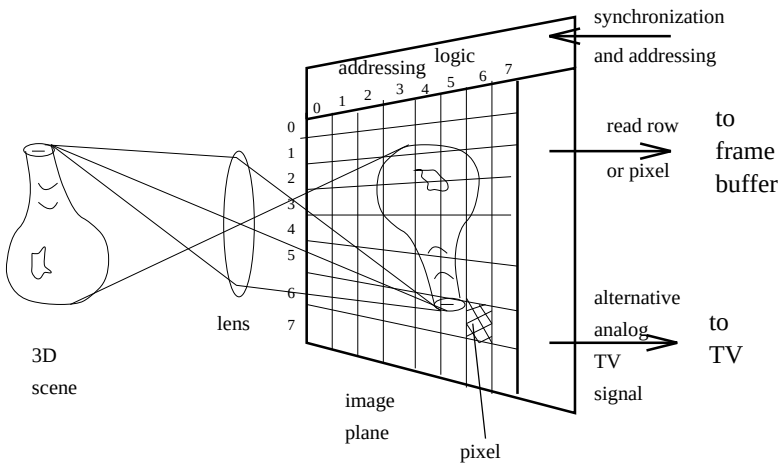

Figure 2.2: A CCD (charge-coupled device) camera imaging a vase; discrete cells convert light energy into electrical charges, which are represented as small numbers when input to a computer.

If the digital image has 500 rows and 500 columns of byte-sized gray values, a memory array of a quarter of a million bytes is obtained. A CCD camera sometimes plugs into a computer board, called a *frame grabber* which contains memory for the image and perhaps control of the camera. New designs now allow for direct digital communication (e.g. using the IEEE 1394 standard). Today major camera manufacturers offer digital cameras that can store a few dozen images in memory within the camera body itself; some contain a floppy disk for this purpose. These images can be input for computer processing at any time. Figure 2.3 sketches an entire computer system with both camera input and graphics output. This is a typical system for an industrial vision task or medical imaging task. It is also typical for *multimedia* computers, which may have an inexpensive camera available to take images for teleconferencing purposes. The role of a *frame buffer* as a high speed

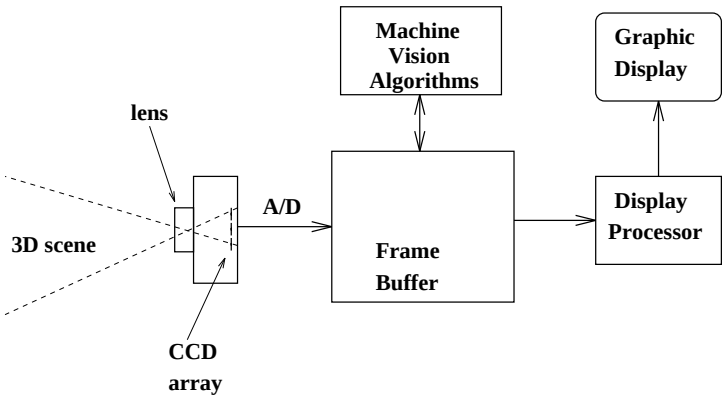

Figure 2.3: Central role of the *frame buffer* in image processing.

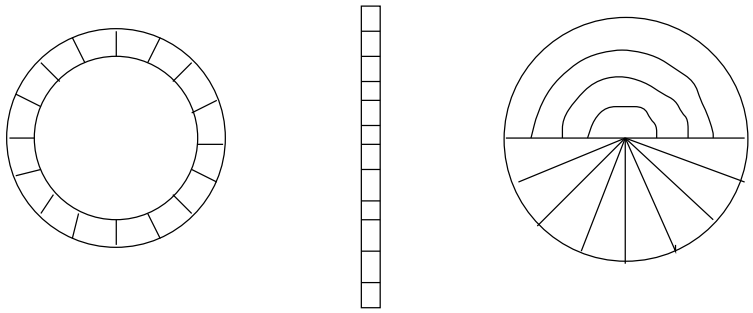

Figure 2.4: Other useful array geometries: (a) circular, (b) linear, (c) "ROSA".

image store is central here: the camera provides an input image which is stored in digital form in the frame buffer after analog to digital conversion where it is available for display to the user and for processing by various computer algorithms. The frame buffer actually may store several images or their derivatives.

A computer program processing a digital image might refer to pixel values as **I[r, c]** or **I[r][c]** where **I** is an array name and **r** and **c** are row and column numbers, respectively. This book uses such notation in the algorithms presented. Some cameras can be set so that they produce a *binary image* – pixels are either 0 or 1 representing dark versus bright, or the reverse. A simple algorithm can produce the same effect by changing all pixels below some *threshold* value t to 0 and all pixels at or above it to 1. An example was given in Chapter 1 where a magnetic resonance image was thresholded to contrast high blood flow versus low blood flow.

Image Formation

The geometry of image formation can be conceptualized as the projection of each point of the 3D scene through the *center of projection* or *lens center* onto the image plane. The intensity at the image point is related to the intensity radiating from the 3D surface point: the actual relationship is complex as we'll later learn. This projection model can be physically justified since a *pin-hole* camera can actually be made by using a camera box with a small hole and no lens at all. A CCD camera usually will employ the same kind of lens as 35mm film cameras used for family photos. A single lens with two convex surfaces is shown in Figure 2.2, but most actual lenses are compound with more than two refracting surfaces. There are two very important points to be made. First, the lens is a light collector: light reaches the image point via an entire cone of rays reaching the lens from the 3D point. Three rays are shown projecting from the top of the vase in Figure 2.2; these determine the extremes of the cone of rays collected by the lens for only the top of the vase. A similar cone of rays exists for all other scene points. Because of geometric imperfections in the lens, different bending of different colors of light, and other phenomena, the cone of rays actually results in a finite or blurred spot on the image plane called the *circle of confusion*. Secondly, the CCD sensor array is constructed of physically discrete units and not infinitesimal points;

thus, each sensor cell integrates the rays received from many neighboring points of a 3D surface. These two effects cause *blurring* of the image and limit its sharpness and the size of the smallest scene details that can be sensed.

CCD arrays are manufactured on chips typically measuring about 1 cm x 1 cm. If the array has 640 x 480 pixels or 512 x 512 pixels, then each pixel has a real width of roughly 0.001 inch. There are other useful ways of placing CCD sensor cells on the image plane (or image line) as shown in Figure 2.4. A linear array can be used in cases where we only need to measure the width of objects or where we may be imaging and inspecting a continuous web of material flowing by the camera. With a linear array, 1000 to 5000 pixels are available in a single row. Such an array can be used in a *push broom* fashion where the linear sensor is moved across the material being scanned as done with a hand held scanner or in highly accurate mechanical scanners, such as flatbed scanners. Currently, many flatbed scanners are available for a few hundred dollars and are used to acquire digital images from color photos or print media. Cylindrical lenses are commonly used to focus a "line" in the real world onto the linear CCD array. The circular array would be handy for inspecting analog dials such as on watches or speedometers: the object is positioned carefully relative to the camera and the circular array is scanned for the image of the needle. The interesting "ROSA" partition shown in Figure 2.4(c) provides a hardware solution to integrating all the light energy falling into either sectors or bands of the circle. It was designed for quantizing the power spectrum of the an image, but might have other simple uses as well. Chip manufacturing technology presents opportunities for implementing other custom designs.

Exercise 1 examination of a CCD camera

If you have access to a CCD camera, obtain permission to explore its construction. Remove the lens and note its construction; does it have a shutter to close off all light, does it have an aperature to change the size of the cone of rays passing through? Is there a means of changing the focal length – e.g. the distance between the lens and CCD? Inspect the CCD array. How large is the active sensing area? Can you see the individual cells – do you need a magnifying glass?

Exercise 2

Suppose that an analog clock is to be read using a CCD camera that stares directly at it. The center of the clock images at the center of a 256 x 256 digital image and the hour hand is twice the width of the minute hand but 0.7 times its length. To locate the images of the hands of the clock we need to scan the pixels of the digital image in a circular fashion. (a) Give a formula for computing $\mathbf{r}(t)$ and $\mathbf{c}(t)$ for pixels $\mathbf{I}[\mathbf{r}, \mathbf{c}]$ on a circle of radius R centered at the image center $\mathbf{I}[\mathbf{256}, \mathbf{256}]$, where t is the angle made between the ray to $\mathbf{I}[\mathbf{r}, \mathbf{c}]$ and the horizontal axis. (b) Is there a problem in controlling t so that a unique sequence of pixels of a *digital circle* is generated? (*c) Do some outside reading in a text on computer graphics and report on a practical method for generating such a digital circle.

Video cameras

Video cameras creating imagery for human consumption record sequences of images at a rate of 30 per second, enabling a representation of object motion over time in addition to the

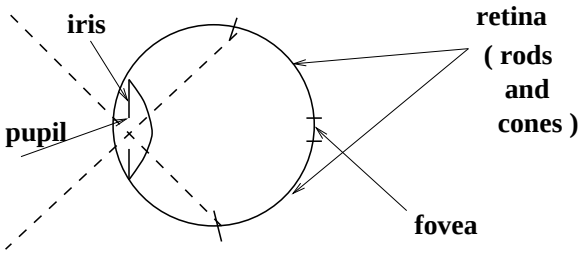

Figure 2.5: Crude sketch of the human eye as camera. (Much more detail can be obtained in the 1985 book by Levine.)

spatial features represented in the single images or *frames*. To provide for smooth human perception, 60 half frames per second are used: these half frames are all odd image rows followed by all even image rows in alternate succession. An audio signal is also encoded. Video cameras creating imagery for machine consumption can record images at whatever rate is practical and need not use the half frame technique.

Frames of a video sequence are separated by markers and some image compression scheme is usually used to reduce the amount of data. The analog TV standards have been carefully designed to satisfy multiple requirements: the most interesting features allow for the same signal to be used for either color or monochrome TVs and to carry sound or text signals as well. The interested reader should consult the related reading and the summary of MPEG encoding given below. We continue here with the notion of digital video being just a sequence of 2D digital images.

CCD camera technology for machine vision has sometimes suffered from display standards designed for human consumption. First, the interlacing of odd/even frames in a video sequence, needed to give a smooth picture to a human makes unnecessary complexity for machine vision. Secondly, many CCD arrays have had pixels with a 4:3 ratio of width to height because most displays for humans have a 4:3 size ratio. Square pixels and a single scale parameter would benefit machine vision. The huge consumer market has driven device construction toward human standards and machine vision developers have had to either adapt or pay more for devices made in limited quantities.

The Human Eye

Crudely speaking, the human eye is a spherical camera with a 20mm focal length lens at the outside focusing the image on the *retina* which is opposite the lens and fixed on the inside of the surface of the sphere (see Figure 2.5). The *iris* controls the amount of light passing through the lens by controlling the size of the *pupil*. Each eye has one hundred million receptor cells – quite a lot compared to a typical CCD array. Moreover, the retina is unevenly populated with sensor cells. An area near the center of the retina, called the *fovea*, has a very dense concentration of color receptors, called *cones*. Away from the center, the density of cones decreases while the density of black-white receptors, the *rods*, increases. The human eye senses three separate intensities for three constituent colors of a single surface spot imaging on the fovea, because the light received from that spot falls

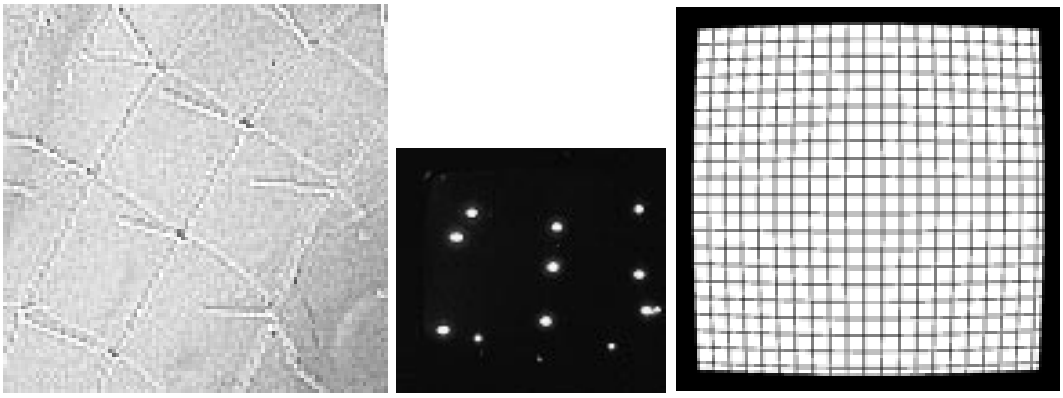

Figure 2.6: Images showing various distortions. (Left) Grey level clipping during A/D conversion occurs at the intersection of some bright stripes; (center) blooming increases the intensity at the neighbors of bright pixels; (right) barrel distortion is often observed when short focal length lenses are used.

on 3 different types of cones. Each type of cone has a special pigment that is sensitive to wavelengths of light in a certain range. One of the most intriguing properties of the human eye-brain is its ability to smoothly perceive a seamless and stable 3D world even though the eyes are constantly moving. These *saccades* of the eye are necessary for proper human visual perception. A significant part of the human brain is engaged in processing visual input. Other characteristics of the human visual system will be discussed at various points in the book: in particular, more details of color perception are given in Chapter 6.

Exercise 3

Assume that a human eyeball is 1 inch in diameter and that 10^8 rods and cones populate a fraction of $1/\pi$ of its inner surface area. What is the average size of area covered by a single receptor? (Remember, however, that foveal receptors are packed much more densely than this average, while peripheral receptors are more sparse.)

2.3 * Problems in Digital Images

Several problems affect the sensing process, some of the most important of which are listed below. Usually, our idealized view given previously is only an approximation to the real physics. The overall effect of the combination of these problems is an image that has some distortion in both its geometry and intensities. Methods for correcting some of these problems are given later in the book; methods for making decisions despite such imperfections are more common, however.

geometric distortion

Geometric distortion is present in several ways in the imaging process. The lens may be imperfect so that the beams of light being collected from a scene surface element are not

bent exactly as intended. Barrel distortion is commonly observed for small focal length lenses; straight lines at the periphery of the scene appear to bow away from the center of the image as shown at the right in Figure 2.6.

scattering

Beams of radiation can be bent or dispersed by the medium through which they pass. Aerial and satellite images are particularly susceptible to such effects, which are caused by water vapor or temperature gradients that give lens-like characteristics to the atmosphere.

blooming

Because discrete detectors, such as CCD cells, are not perfectly insulated from each other, charge collected at one cell can leak into a neighboring cell. The term *blooming* arises from the phenomena where such leakage spreads out from a very bright region on the image plane, resulting in a bright "flower" in the image that is larger than it actually should be as shown in Figure 2.6(center).

CCD variations

Due to imperfections in manufacturing, there may be variations in the responses of the different cells to identical light intensity. For precise interpretation of intensity, it may be necessary to determine a full array of scale factors $s[r,c]$ and shifts $t[r,c]$, one for each pixel, by calibration with uniform illumination so that intensity can be restored as $I_2[r,c] = s[r,c]I_1[r,c] + t[r,c]$. In an extreme case, the CCD array may have some *dead cells* which give no response at all. Such defects can be detected by inspection: one software remedy is to assign the response of a dead cell to be the average response of the neighbors.

clipping or wrap-around

In the analog to digital conversion, a very high intensity may be clipped off to a maximum value, or, its high order bits may be lost, causing the value to be wrapped-around into some encoding for a lower intensity. The result of wrap-around is seen in a grey-scale image as a bright region with a darker core; in a color image it can result in a noticeable change in color. The image at the left in Figure 2.6 shows wrap-around: some intersections of bright lines result in pixels darker than those for either line.

chromatic distortion

Different wavelengths of light are bent differently by a lens (the *index of refraction* of the lens varies with wavelength). As a result, energy in different wavelengths of light **from the same scene spot** may actually image a few pixels apart on the detector. For example, the image of a very sharp black-white boundary in the periphery of the scene may result in a *ramp* of intensity change spread over several pixels in the image.

quantization effects

The digitization process collects a sample of intensity from a discrete area of the scene and maps it to one of a discrete set of grey values and thus is susceptible to both mixing and

rounding problems. These are addressed in more detail in the next section.

2.4 Picture Functions and Digital Images

We now discuss some concepts and notation important for both the theory and programming of image processing operations.

Types of images

In computing with images, it is convenient to work with both the concepts of *analog image* and *digital image*. The picture function is a mathematical model that is often used in analysis where it is fruitful to consider the image as a function of two variables. All of functional analysis is then available for analyzing images. The digital image is merely a 2D rectangular array of discrete values. Both image space and intensity range are quantized into a discrete set of values, permitting the image to be stored in a 2D computer memory structure. It is common to record intensity as an 8-bit (1-byte) number which allows values of 0 to 255. 256 different levels is usually all the precision available from the sensor and also is usually enough to satisfy the consumer. And, bytes are convenient for computers. For example, an image might be declared in a C program as "**char I[512][512];**". Each pixel of a color image would require 3 such values. In some medical applications, 10-bit encoding is used, allowing 1024 different intensity values, which approaches the limit of humans in discerning them.

The following definitions are intended to clarify important concepts and also to establish notation used throughout this book. We begin with an ideal notion of an analog image created by an ideal optical system, which we assume to have infinite precision. Digital images are formed by *sampling* this analog image at discrete locations and representing the intensity at a location as a discrete value. All real images are affected by physical processes that limit precision in both position and intensity.

1 Definition *An* **analog image** *is a 2D image* **F(x,y)** *which has infinite precision in spatial parameters x and y and infinite precision in intensity at each spatial point* **(x, y)**.

2 Definition *A* **digital image** *is a 2D image* **I[r, c]** *represented by a discrete 2D array of intensity samples, each of which is represented using a limited precision.*

The mathematical model of an image as a function of two real spatial parameters is enormously useful in both describing images and defining operations on them. Figure 2.7(d) shows how the pixels of an image are samples of a continuous image taken at various points $[x, y]$ of the image plane. If there are M samples in the X-direction across a distance of w, then the x-spacing Δx between pixels is w/M. The formula relating the center point of a pixel to the array cell containing the intensity sample is given in the figure at the right.

3 Definition *A* **picture function** *is a mathematical representation $f(x, y)$ of a picture as a function of two spatial variables x and y. x and y are real values defining points of the picture and $f(x, y)$ is usually also a real value defining the intensity of the picture at point (x, y).*

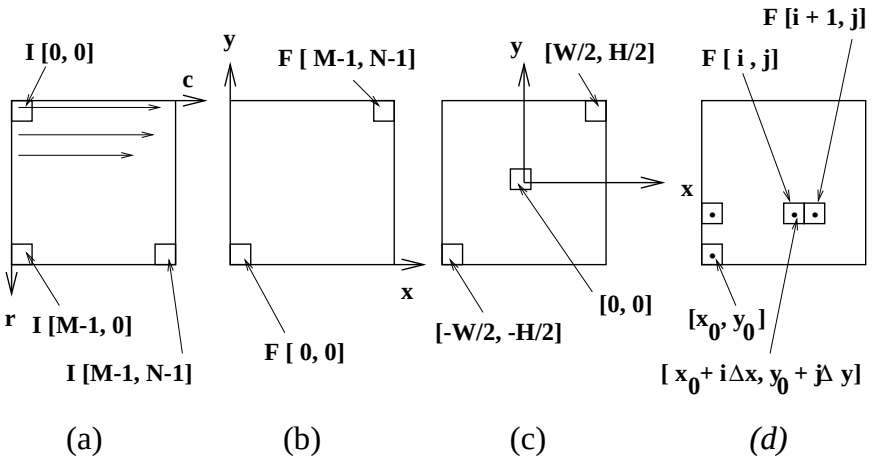

Figure 2.7: Different coordinate systems used for images: (a) *raster oriented* uses row and column coordinates starting at $[0,0]$ from the top left; (b) Cartesian coordinate frame with $[0,0]$ at the lower left; (c) Cartesian coordinate frame with $[0,0]$ at the image center. (d) Relationship of pixel center point $[x,y]$ to area element sampled in array element $I[i,j]$.

4 Definition *A* **grey scale image** *is a monochrome digital image* **I[r, c]** *with one intensity value per pixel.*

5 Definition *A* **multispectral image** *is a 2D image* **M[x,y]** *which has a vector of values at each spatial point or pixel. If the image is actually a color image, then the vector has 3 elements.*

6 Definition *A* **binary image** *is a digital image with all pixel values 0 or 1.*

7 Definition *A* **labeled image** *is a digital image* **L[r, c]** *whose pixel values are* symbols *from a finite alphabet. The symbol value of a pixel denotes the outcome of some decision made for that pixel. Related concepts are* **thematic image** *and* **pseudo-colored image**.

A coordinate system must be used to address individual pixels of an image; to operate on it in a computer program, to refer to it in a mathematical formula, or to address it relative to device coordinates. Different systems used in this book and elsewhere are shown in Figure 2.7. Unfortunately, different computer tools often use different systems and the user will need to get accustomed to them. Fortunately, concepts are not tied to a coordinate system. In this book, concepts are usually discussed using a Cartesian coordinate system consistent with mathematics texts while image processing algorithms usually use raster coordinates.

Image Quantization and Spatial Measurement

Each pixel of a digital image repesents a sample of some elemental region of the real image as is shown in Figure 2.2. If the pixel is projected from the image plane back out to the source material in the scene, then the size of that scene element is the *nominal resolution*

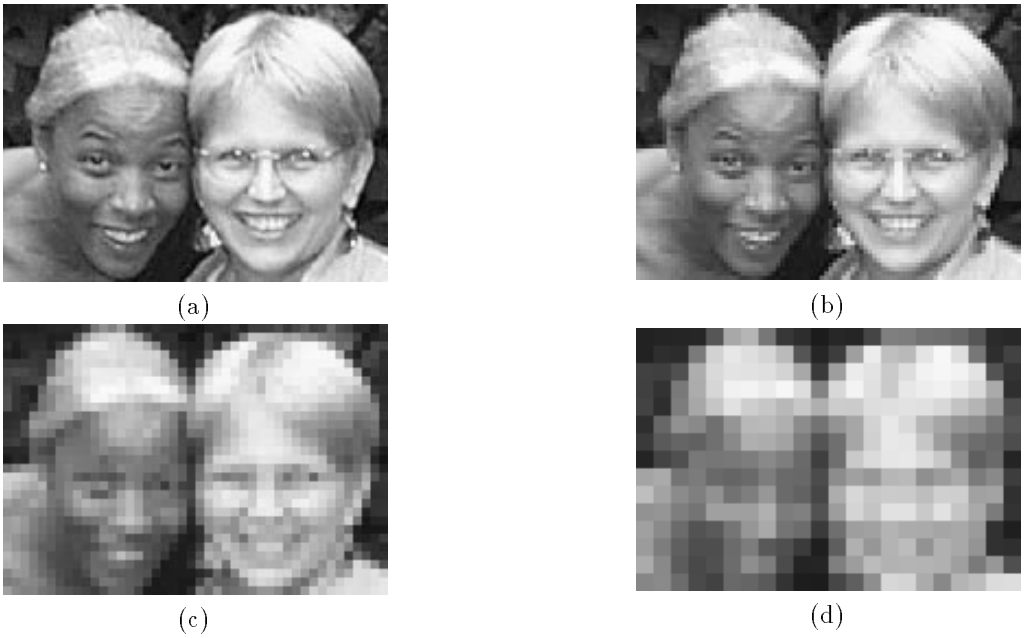

Figure 2.8: Four digital images of two faces; (a) 127 rows of 176 columns; (b) (126x176) created by averaging each 2×2 neighborhood of (a) and replicating the average four times to produce a 2×2 average block; (c) (124x176) created in same manner from (b); (d) (120x176) created in same manner from (c). Effective nominal resolutions are (127x176), (63x88), (31x44), (15x22) respectively. (Try looking at the blocky images by squinting; it usually helps by blurring the annoying sharp boundaries of the squares.) Photo courtesy of Frank Biocca.

of the sensor. For example, if a 10 inch square sheet of paper is imaged to form a 500 x 500 digital image, then the nominal resolution of the sensor is 0.02 inches. This concept may not make sense if the scene has a lot of depth variation, since the nominal resolution will vary with depth and surface orientation. The *field of view* of an imaging sensor is a measure of how much of the scene it can see. The *resolution* of a sensor is related to its precision in making spatial measurements or in detecting fine features. (With careful use, and some model information, a 500 x 500 pixel image can be used to make measurements to an accuracy of 1 part in 5000, which is called *subpixel resolution*.)

8 Definition *The* **nominal resolution** *of a CCD sensor is the size of the scene element that images to a single pixel on the image plane.*

9 Definition *The term* **resolution** *refers to the precision of the sensor in making measurements, but is formally defined in different ways. If defined in real world terms, it may just be the nominal resolution, as in "the resolution of this scanner is one meter on the ground" or it may be in the number of line pairs per millimeter that can be "resolved" or distinguished in the sensed image. A totally different concept is the number of pixels available*

– "the camera has a resolution of 640 by 480 pixels". This later definition has an advantage in that it states into how many parts the field of view can be divided, which relates to both the capability to make precise measurements and to cover a certain region of a scene. If precision of measurement is a fraction of the nominal resolution, this is called **subpixel resolution**.

Figure 2.8 shows four images of the same face to emphasize resolution effects: humans can recognize a familiar face using 64x64 resolution, and maybe using 32x32 resolution, but 16x16 is insufficient. In solving a problem using computer vision, the implementor should use an appropriate resolution; too little resolution will produce poor recognition or imprecise measurements while too much will unnecessarily slow down algorithms and waste memory.

10 DEFINITION *The* **field of view** *of a sensor (***FOV***) is the size of the scene that it can sense, for example 10 inches by 10 inches. Since this may vary with depth, it may be more meaningful to use* **angular field of view**, *such as 55 degrees by 40 degrees.*

Since a pixel in an image measures an area in the real world and not a point, its value often is determined by a mixture of different materials. For example, consider a satellite image where each pixel samples from a spot of the earth 10m x 10m. Clearly, that pixel value may be a sample of water, soil, and vegetation combined. The problem appears in a severe form when binary images are formed. Reconsider the above example of imaging a sheet of paper with 10 characters per inch. Many image pixels will overlap a character boundary and hence receive a mixture of higher intensity from the background and lower intensity from the character; the net result being a value in between background and character that could be set to either 0 or 1. Whichever value it is, it is partly incorrect!

Figure 2.9 gives details of quantization problems. Assume that the 2D scene is a 10x10 array of black (brightness 0) and white (brightness 8) tiles as shown at the left in the figure. The tiles form patterns that are 2 bright spots and two bright lines of different widths. If the image of the scene falls on a 5x5 CCD array such that each 2x2 set of adjacent tiles falls precisely on one CCD element the result is the digital image shown in Figure 2.9(b). The top left CCD element senses intensity $2 = (0 + 0 + 0 + 8)/4$ which is the average intensity from four tiles. The set of four bright tiles at the top right fall on two CCD elements, each of which integrates the intensity from two bright and two dark tiles. The single row of bright tiles of intensity 8 images as a row of CCD elements sensing intensity 4, while the double row images as two rows of intensity 4; however, the two lines in the scene are blended together in the image. If the image is thresholded at $t = 3$, then a bright pattern consisting of one tile will be lost from the image and the three other features will all fuse into one region! If the camera is displaced by an amount equivalent to one tile in both the horizontal and vertical direction, then the image shown in Figure 2.9(d) results. The shape of the 4-tile bright spot is distorted in (d) a different manner than in (b) and the two bright lines in the scene result in a "ramp" in (d) as opposed to the constant grey region of (b); moreover, (d) shows three object regions whereas (b) shows two. Figure 2.9 shows that the images of scene features that are nearly the size of one pixel are unstable.

Figure 2.9, shows how *spatial quantization effects* impose limits on measurement accuracy and detectability. Small features can be missed or fused and even when larger features are detected, their spatial extent might be poorly represented. Note how the bright set

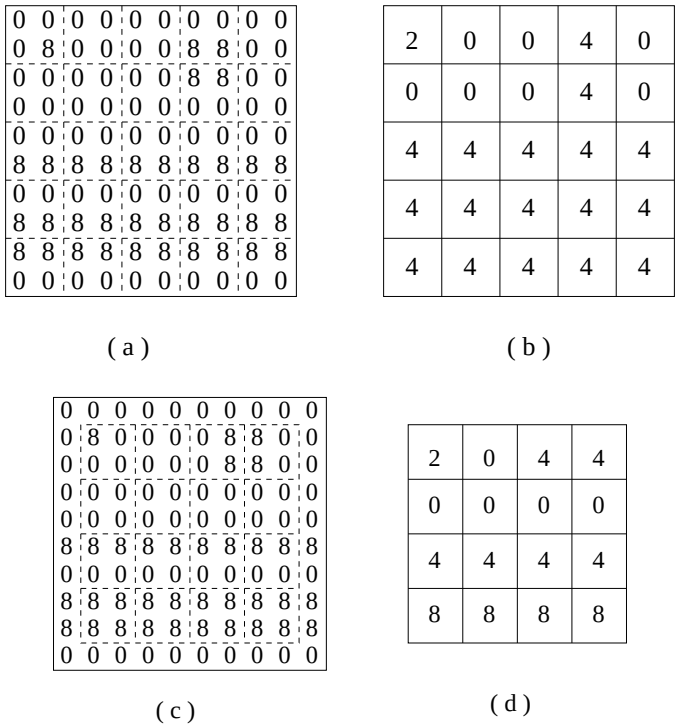

Figure 2.9: (a) 10×10 field of tiles of brightness 0 or 8. (b) Intensities recorded in a 5x5 image of precisely the brightness field at the left where each pixel senses the average brightness of a 2×2 neighborhood of tiles. (c) Image sensed by shifted camera "one tile down and one tile to the right". Note that the quantized brightness values depend on both the actual pixel size and position relative to the brightness field. (d) Intensities recorded from the shifted camera in the same manner as in (b). Interpretation of the actual scene features will be problematic with either image (b) or (d).

of four tiles images as either a vertical or horizontal pair of CCD elements of intensity 4. Perhaps we should expect an error as bad as 0.5 pixels in the placement of a boundary due to rounding of a *mixed pixel* when a binary image is created by thresholding; this implies a one pixel expected error in a measurement made across two boundaries. Moreover, *if we expect to detect certain features in a binary image, then we must make sure that their image size is at least two pixels in diameter: this includes gaps between objects.* Consider a "period" ending a sentence in a FAX whose image is one pixel in diameter but falls exactly centered at the point where 4 CCD cells meet: each of the 4 pixels will be mixed with more background than character and it is likely that the character will be lost when a binary image is formed!

11 DEFINITION *A* **mixed pixel** *is an image pixel whose intensity represents a sample from a mixture of material types in the real world.*

Exercise 4 On variation of area

Consider a dark rectangle on a white paper that is imaged such that the rectangle measures exactly 5.9 x 8.1 pixels on the real image. A binary image is to be produced such that pixel values are 0 or 1 depending upon whether the pixel sees more object or more background. Allow the rectangle to translate with its sides parallel to the CCD rows and columns. What is the smallest area in pixels in the binary image output? What is the largest area?

Exercise 5 On loss of thin features

Consider two bright parallel lines of conductor on a printed circuit board. The width of each line is 0.8 pixels on the image plane. Is there a situation where one line and not the other would disappear in a binary image created as in the exercise above? Explain.

In Chapter 13, the thin lens equation from optics is reviewed and is studied with respect to how it relates camera resolution, image blur, and depth of field: the interested reader will be able to understand that section at this point. Having taken some care to discuss the characteristics of sensing and the notions of resolution and mixed pixels, we now have enough background to begin working on certain 2D machine vision applications. We might want to find certain objects using a microscope, inspect a PC board, or recognize the shadow of a backlit 3D object. The imaging environment must be engineered so that the features that must be seen are of the proper size in the image. Assuming that there is no significant 3D character remaining in the image after the scaling from world to image is considered, the images can then be analyzed using the 2D methods of the next several chapters.

2.5 * Digital Image Formats

Use of digital images is widespread in communication, databases, and machine vision and standard formats have been developed so that different hardware and software can share data. Figure 2.10 sketches this situation. Unfortunately, there are dozens of different formats still in use. A few of the most important ones are briefly discussed in this section. A *raw image* may be just a stream of bytes encoding the image pixels in row-by-row order,

Exercise 6 sensing paper money denominations

Consider the design of a sensor for a vending machine that takes U.S. paper money of denominations \$1, \$5, \$10, and \$20. You only need to create a representation for the recognizer to use; you need not design a recognition algorithm, nor should you be concerned about detecting counterfeit bills. (Be sure to obtain some samples before answering.) Assume that a linear CCD array must be used to digitize a bill as it enters the machine. (a) What kind of lens and what kind of illumination should be used? (b) How many pixels are needed in the linear array? Explain.

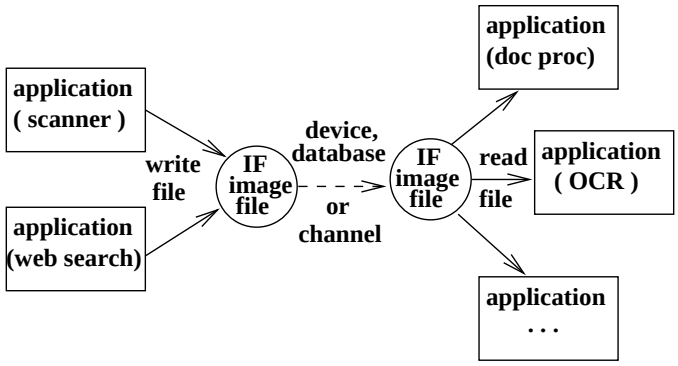

Figure 2.10: Many devices or application programs create, consume, or convert image data. Standard format image files (**IF**s) are needed to do this productively for a family of devices and programs.

called *raster order*, perhaps with line-feeds separating rows. Information such as image type, size, time taken, and creation method is not part of a raw image. Such information might be handwritten on a tape label or in someone's research notebook – this is inadequate. (One project, in which one author took part, videotaped a bar code before videotaping images from the experiment. The computer program would then process the bar code to obtain overall non-image information about the experimental treatment.) Most recently developed standard formats contain a header with non-image information necessary to label the data and to decode it.

Several formats originated with companies creating image processing or graphics tools; in some cases but not in others, public documentation and conversion software are available. The details provided below should provide the reader with practical information for handling computer images. Although the details are changing rapidly with technology, there are several general concepts contained in this section that should endure.

Image File Header

A *file header* is needed to make an image file self-describing so that image processing tools can work with them. The header should contain the image dimensions, type, date of creation, and some kind of title. It may also contain a color table or coding table to be used to interpret pixel values. A nice feature not often available is a *history section* containing notes on how the image was created and processed.

Image Data

Some formats can handle only limited types of images, such as binary and monochrome; however, those surviving today have continued to grow to include more image types and features. Pixel size and image size limits typically differ between different file formats. Several formats can handle a sequence of frames. *Multimedia* formats are evolving and include image data along with text, graphics, music, etc.

Data Compression

Many formats provide for *compression* of the image data so that all pixel values are not directly encoded. Image compression can reduce the size of an image to 30% or even 3% of its raw size depending on the quality required and method used. Compression can be *lossless* or *lossy*. With lossless compression, the original image can be recovered exactly. With lossy compression, the pixel representations cannot be recovered exactly: sometimes a loss of quality is perceived, but not always. To implement compression, the image file must include some overhead information about the compression method and parameters. Most digital images are very different from symbolic digital information – loss or change of a few bits of digital image data will have little or no effect on its consumer, regardless of whether it is a human or machine. The situation is quite different for most other computer files; for example, changing a single bit in an employee record could change the salary field by $8192 or the apartment address from 'A' to 'B'. Image compression is an exciting area that spans the gamut from signal processing to object recognition. Image compression is discussed at several points of this textbook, but is not systematically treated.

12 Definition *An image compression method is* **lossless** *if a decompression method exists to precisely recover (every bit of) the original image representation. Otherwise, the compression method is* **lossy**.

Commonly used Formats

Many of the images in this book passed through multiple formats. Some images were received from a colleague or retrieved from an image database in GIF, JPG or even PS format. Some were scanned from photos and their original digital format was GIF or TIFF. Simple image processing might have been done using the image tool xv and more complex operations were done with hips tools or with special C or C++ programs. Some of the most commonly used formats are briefly described below. **Image/Graphics file formats are still evolving** with a trend for each to be more inclusive. The reader should be aware that some of the details given below will have to be updated by consulting the latest reference document.

run-coded binary images

Run-coding is an efficient coding scheme for binary or labeled images: not only does it reduce memory space, but it can also speed up image operations, such as set operations. Run-coding works well when there is a lot of redundancy in pixels along the image rows.

```
                  00000000001111111111222222222233333333334444444444
Column c     :    01234567890123456789012345678901234567890123456789

Image Row r :     00000000111110000000000001110000000111111111100000

Run-code A   :    8(0)5(1)12(0)3(1)7(0)9(1)5(0)

Run-code B   :    (8,12)(25,27)(35,43)
```

Figure 2.11: Runcoding encodes the runs of consecutive 0 or 1 values, and for some domains, yields an efficiently compressed image.

Assume a binary image; for each image row, we could record the **number** of 0's followed by the number of 1's alternating across the entire row. Figure 2.11A gives an example. Run-code B of the figure shows a more compact encoding of just the 1-runs from which we can still recover the original row. We will use such encodings for some algorithms in this book. Run-coding is often used for compression within standard file formats.

PGM: Portable Grey Map

One of the simplest file formats for storing and exchanging image data is the **PBM** or "**P**ortable **B**it **M**ap family of formats (PBM/PGM,PPM). The image header and pixel information are encoded in ASCII. The image file representing an image of 8 rows of 16 columns with maximum grey value of 192 is shown in Figure 2.12. Two graphic renderings are also shown, each is the output of image conversion tools applied to the original text input. The image at the lower left was made by replicating the pixels to make a larger image of 32 rows of 64 columns each; the image at the lower right was made by first converting to JPG format with lossy compression. The first entry of the PGM file is the *Magic Value*,"P2" in our example, indicating how the image information is coded (ASCII grey levels in our example). Binary, rather than ASCII pixel coding is available for large pictures. (The magic number for binary is "P4").

Exercise 7 Creating a PPM picture

A color image can be coded in PBM format by using the magic value "P3" and three (R,G,B) intensity values for each pixel, similar to the coded monochrome "P2" file shown in Figure 2.12. Using your editor, create a file bullseye.ppm encoding 3 coincentric circular regions of different colors. For each pixel, the three color values follow in immediate succession, rather than encoding three separate monochrome images as is done in some other formats. Display your picture using an image tool or web browser.

```
P2
# sample small picture 8 rows of 16 columns, max grey value of 192
# making an image of the word "Hi".
  16 8   192

 64 64  64  64  64  64  64  64  64 64 64  64  64 64 64 64
 64 64 128 128  64  64  64 128 128 64 64 192 192 64 64 64
 64 64 128 128  64  64  64 128 128 64 64 192 192 64 64 64
 64 64 128 128 128 128 128 128 128 64 64  64  64 64 64 64
 64 64 128 128 128 128 128 128 128 64 64 128 128 64 64 64
 64 64 128 128  64  64  64 128 128 64 64 128 128 64 64 64
 64 64 128 128  64  64  64 128 128 64 64 128 128 64 64 64
 64 64  64  64  64  64  64  64  64 64 64  64  64 64 64 64
```

Figure 2.12: Text (ASCII) file representing an image of the word "Hi"; 64 is the background level, 128 is the level of "H" and the lower part of "i', and 192 is the level of the dot of the "i". At the lower left is a printed picture made from the above text file using image format conversion tools. At the bottom right is an image made using a lossy compression algorithm.

GIF image file format

The **G**raphics **I**nterchange **F**ormat (GIF) originated from CompuServe, Inc. and has been used to encode a huge number of images on the WWW or in current databases. GIF files are relatively easy to work with, but cannot be used for high-precision color, since only 8-bits are used to encode color. The 256 color values available are typically sufficient for computer displayed images; a more compact 16-color option can also be used. LZW non-lossy compression is available.

TIFF image file format

Originated by Aldus Corp., TIFF or TIF is very general and very complex. It is used on all popular platforms and is often the format used by scanners. **T**ag **I**mage **F**ile **F**ormat supports multiple images with 1 to 24 bits of color per pixel. Options are available for either lossy or lossless compression.

JPEG format for still photos

JPEG (JFIF/JFI/JPG) is a more recent standard from the Joint Photographic Experts Group; the major purpose was to provide for practical compression of high-quality color still images. The JPEG coding scheme is stream-oriented and allows for real-time hardware for coding and decoding. An image can have up to 64k x 64k pixels of 24 bits each, although there is only one image per file. The header can contain a thumbnail image of up to 64k uncompressed bytes. JPEG is indepenent of the color coding system, a major advantage. More details on color systems are given in Chapter 6. To achieve high compression, a flexible, but complex lossy coding scheme is used which often can compress a high quality image

20:1 without noticable degradation. The compression works well when the image has large regions of nearly constant color and when high frequency variation in regions of detail is not important to the consumer. (JPEG has a little used lossless compression option, which might achieve 2:1 compression using *predictive coding*.) The compression scheme uses the *discrete cosine transformation*, which is discussed in Chapter 5, followed by *Huffman coding*, which is not treated in this book. JPEG is **not** designed for video.

Exercise 8

Locate an image viewing toolset on your computer system. (These might be available just by clicking on your image file icon.) Use one image of a face and one of a landscape; both, should originally be of high quality, say 800 x 600 color pixels, from a flatbed scanner or digital camera. Convert the image among different formats – GIF,TIFF,JPEG etc. Record the size of the encoded image files in bytes and note the quality of the image; consider the overall scene plus small details.

Exercise 9 * JPEG study

(a) Research the JPEG compression scheme for 8×8 image blocks. (b) Implement and test the DCT scheme in a lossless manner (except for possible roundoff error). (c) Create a lossy compression using some existing image tool. (d) Using the 64 coefficients from lossy compression, regenerate an 8×8 image and compare its pixel values with those of the original 8×8 image.

PostScript

The family of formats BDF/PDL/EPS store image data using printable ASCII characters and are often used with X11 graphics displays and printers. "PDL" is a page description language and "EPS" is encapsulated postscript (originally from Adobe), which is commonly used to contain graphics or images to be inserted into a larger document. Pixels values are encoded via 7-bit ASCII codes, so these files can be examined and changed by a text editor. 75 to 3000 dots per inch of grey scale or color can be represented and newer versions include JPEG compression. A PDL header contains the bounding box of the image on the page where it is to appear. Most of the images in this book have been included as EPS files.

MPEG format for video

MPEG (MPG/MPEG-1/MPEG-2) is a stream-oriented encoding scheme for video, audio, text and graphics. "MPEG" stands for **M**otion **P**icture **E**xperts **G**roup, an international group of representatives from industry and governments. The MPEG family of standards is currently evolving rapidly along with the technology of computers and communication. MPEG-1 is primarily designed for multimedia systems and provides for a data rate of 0.25 Mbits per second of compressed audio and 1.25 Mbits of compressed video. These rates are suitable for multimedia for popular personal computers, but is too low for high-quality TV. The MPEG-2 standard provides for up to 15 Mbits per second data rates to handle high definition TV rates. The compression scheme takes advantage of both spatial redundancy, as used in JPEG, and temporal redundancy and generally provides a useful compression ratio of 25 to 1, with 200 to 1 ratios possible. *Temporal redundancy* essentially means that

Image File Format	No. Bytes "Hi"	No. Bytes "Cars"
PGM	595	509,123
GIF	192	138,267
TIF	918	171,430
PS	1591	345,387
HIPS	700	160,783
JPG (lossless)	684	49,160
JPG (lossy)	619	29,500

Table 2.1: File sizes (in bytes) for the "same" image encoded in different formats: 8×16 grey level "HI" image shown in Figure 2.12 and 347×489 color "Cars" image shown in Figure 2.13.

many regions do not change much from one frame to the next and an encoding scheme can just encode changes and even predict frames from frames before and after in the video sequence. (Future versions of MPEG will have codes for recognized objects and program code to generate their images.) Media quality is determined at the time of encoding. Motion JPEG is a hybrid scheme which just applies JPEG compression to video single frames and does not take advantage of temporal redundancy. While encoding and decoding is simplified using Motion JPEG, compression is not as good, so memory useage and transmission will be poorer than with MPEG. Use of motion vectors by MPEG for compression of video is described in Chapter 9.

Comparison of Formats

Table 2.1 compares some popular image formats in terms of storage size. The left columns of the table apply to the tiny 8×16 greyscale picture "Hi" whereas the right column applies to a 347×489 color image. It is possible to obtain different size pictures for the "same" image by using different sequences of format conversions. For example, the CARS.TIF file output from the scanner was 509,253 bytes, whereas a conversion to a GIF file with only 256 colors required 138,267 bytes, and a TIF file derived from that required 171,430 bytes. This final TIF file had fewer bits in the color codes, but appeared to be qualitatively the same viewed on the CRT. The JPEG file one third its size also displayed the same. While the lossy JPEG is a clear winner in terms of space, this is at a cost of decoding complexity which may require hardware for real-time performance.

2.6 Richness and Problems of Real Imagery

A brief walk with open eyes and mind will confirm what the artist already knows about the richness of the natural visual world. This richness enhances human experience but causes problems for machine vision. (See Figure 2.13 for example.) The intensity or color at an image point depends in complex ways on material, geometry, and lighting; not only is the type of material important, but so is its orientation relative to the sensor, light sources, and other objects. There are specularities on shiny surfaces, shadows, mutual reflection, and transparent materials, for example. For recognition of many surfaces or objects, color

may be of little importance relative to shape or texture, characteristics that depend upon many pixels, not just one. Cases where we have little control over the environment, such as monitoring traffic patterns, can be interesting and difficult.

Problems still remain even in well-engineered industrial environments or TV studios. As we shall see in Chapter 6, reflection from a shiny metal cylinder illuminated by a point source can vary in intensity over a range of 100,000 to 1 and most sensors cannot handle such a dynamic range. Sunlight or artificial light can heat surfaces, causing them to radiate differently over time, brightening CCD images with increased infrared or leaving shadows of airplanes on a runway after their departure. Controlled monochrome laser light can greatly assist some imaging operations, but it can also be totally absorbed by certain surfaces or be dominated by secondary reflections on others.

In many applications of automation, problems can be solved by engineering. Irrelevant bandwidths of light can be filtered out; for example, bruises in dark red cherries can be seen more clearly if a filter is used to allow only infrared light to pass. Moving objects that would cause a blurred image under steady illumination can be illuminated by a *strobe light* for a very short period of time, so that they appear still in an image formed by a highly sensitive detector. Use of *structured light* can make surface measurement and inspection much easier: for example, turbine blades can be illuminated by precise alternating stripes of red and green light, so that many surface defects appear as obvious breaks in the smoothness of the stripes in the 2D image. We will return to these methods at various points of the text.

2.7 3D Structure from 2D Images

The human vision system perceives the structure of the 3D world by integrating several different cues. Here we give only a qualitative description. Cognitive psychologist J.J. Gibson outlined quantitative models for many of these cues. Implementation and demonstration of these models was intensely pursued by computer vision researchers in the 1980's, and several of the quantitative models will be discussed later in the book.

The imaging process records complex relationships between the 3D structure of the world and the 2D structure of the image. Assume the perspective projection model that was described with Figure 2.2 and refer to Figure 2.13. *Interposition* is perhaps the most important depth cue: objects that are closer occlude parts of objects that are farther away: recognition of occlusions gives relative depth. A person seen within the region of a wall is clearly closer to the sensor than the wall. A person recognized behind a car is farther away than the car. *Relative size* is also an important cue. The image of a car 20 meters away will be much smaller than the image of a car 10 meters away, even though the far car might be a larger car. Cars appear to us to be both tiny and moving slowly in the distance; our experience has taught us how to relate the size and speed to the distance. As we walk down the railroad track, the rails appear to meet at a point in the distance (the *vanishing point*), although we know that they must maintain the same separation in 3D. A door that is open into our room images on our retina as a trapezoid and not the rectangle we know it to be. The far edge of the door appears shorter than the near edge; this is the *forshortening* effect of perspective projection and conveys information about the 3D orientation of the door. A related cue is *texture gradient*. The texture of surfaces changes with both the distance from

Figure 2.13: A complex scene with many kinds of depth cues for human perception.

the viewer and the surface orientation. In the park, we can see individual blades of grass or maple leaves up close, while far away we see only green color. The change of image texture due to perspective viewing of a surface receding in the distance is called the *texture gradient*. Chapter 12 gives much more discussion of the issues just mentioned.

Exercise 10 observe as an artist

Consciously make observations in two different environments and sketch some of the cues discussed above. For example, try a busy cafeteria, or a city street corner observed from a height of a few floors, or a spot in a woods.

2.8 Five Frames of Reference

Reference frames are needed in order to do either qualitative or quantitative analysis of 3D scenes. Five frames of reference are needed for general problems in 3D scene analysis, such as controlling operations in a work cell with robots and sensors or providing a virtual 3D environment for human interaction. Several of these frames are not only important for robotics, but are also important to psychologists and the understanding of human spatial perception. The five types of frames are illustrated in Figure 2.14; actually, 6 reference frames are shown since there are two different objects in the scene, a block and a pyramid, each with its own reference frame. In all of these coordinate frames, coordinates are real numbers along continuous axes, except for image coordinates, which are integer subscripts of the pixel array. For the examples in the discussion below you should also imagine an analogous situation where the camera is a TV camera covering a baseball game and objects in the scene are the players, bases, balls, bats, etc.

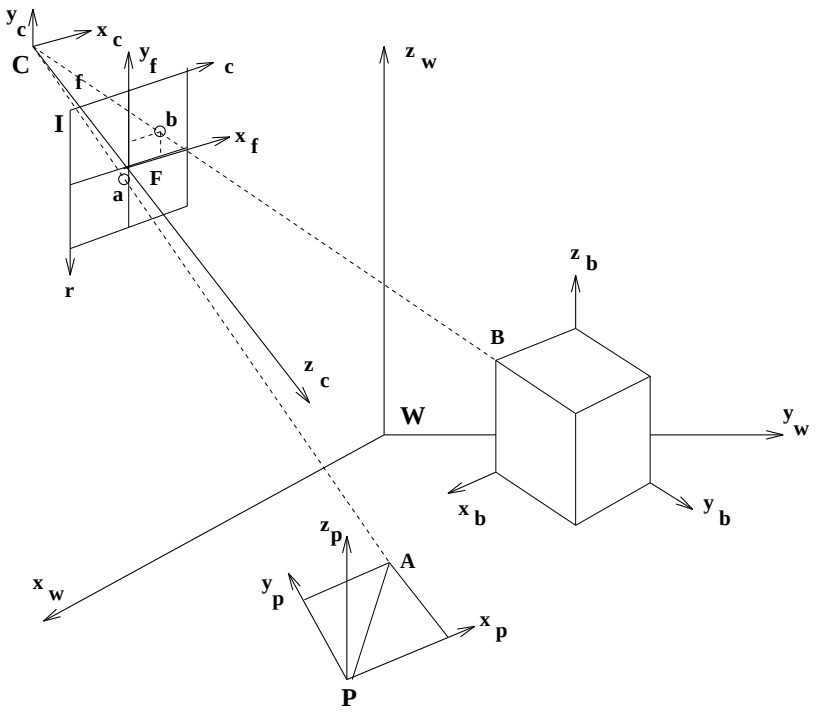

Figure 2.14: Five coordinate frames needed for 3D scene analysis: world **W**, object **O** (for pyramid $\mathbf{O_p}$ or block $\mathbf{O_b}$), camera **C**, real image **F** and pixel image **I**.

Pixel Coordinate Frame I

In the pixel array, each point has integer pixel coordinates. In Figure 2.14, the image of the tip of the pyramid **A** falls within pixel $\mathbf{a} = [a_r, a_c]$ where a_r and a_c are integer row and column, respectively. Many things about a scene can be determined by analysis of the image in terms of only pixel rows and columns. For example, if a pick-and-place robot or other transfer mechanism always delivered a block (or box of laundry detergent) roughly frontal to the camera, then the markings on its frontal surface could be inspected using only the image as a matrix of rows and columns of pixels. In the baseball game analogy, using only the image, one could determine if a batter was using a black bat. Using only image I, however, and no other information, we cannot determine which object is actually larger in 3D or whether or not objects are on a collision course.

Object Coordinate Frame O

An object coordinate frame is used to model ideal objects in both computer graphics and computer vision. For example, Figure 2.14 shows two object coordinate frames, one for a block $\mathbf{O_b}$ and one for a pyramid $\mathbf{O_p}$. The coordinates of 3D corner point **B** relative to the object coordinate frame are $[x_b, 0, z_b]$. These coordinates remain the same, regardless of how the block is posed relative to the world or workspace coordinate frame **W**. The object coordinate frame is needed to inspect an object; for example, to check if a particular hole is in proper position relative to other holes or corners.

Camera Coordinate Frame C

The camera coordinate frame C is often needed for an "egocentric" (cameracentric) view; for example, to represent whether or not an object is just in front of the sensor, moving away, etc. A ball whose image continues to enlarge in the center of your retina is likely to hit you. A seeing robot or human are both object and sensor[s] so their object and sensor coordinate systems may be almost, but not exactly, the same. (Did you ever run into a doorway even though it looked as if you'd pass through without contact?) Computer graphics systems allow the user to select different camera views of the 3D scene being viewed. A play at first base might be better viewed by a camera pointing there.

Real Image Coordinate Frame F

Camera coordinates are real numbers, usually in the same units as the world coordinates – say, inches or mm – *including the depth coordinate* z_c. 3D points project to the real image plane at coordinates $[x_f, y_f, f]$ where f is the focal length. x_f and y_f are not subscripts of pixels in the image array, but are related to the pixel size and pixel position of the optical axis in the image. In Figure 2.14, both coordinates of real image point **a** relative to frame **F** are negative. Frame **F** "contains" the picture function that is digitized to form the digital image in the pixel array **I**.

World Coordinate Frame W

The coordinate frame **W** is needed to relate objects in 3D; for example, to determine whether or not a runner is far off a base or if the runner and second baseman will collide. In a robotics cell or virtual environment, actuators and sensors often communicate via world coordinates; for example, the image sensor tells the robot where to pick up a bolt and in which hole to insert it.

Geometrical and mathematical relationships among these coordinate systems will be very important later in the book. For the next several chapters, however, we process only the information contained in the pixel array under the assumption that there is a straightforward correspondence to the real world. The reader who must work with the algebra of the perspective transformation or its scaling effect will be able at this point to go forward to Chapter 12 to study the perspective imaging model.

2.9 * Other Types of Sensors

This section includes the description of several more sensors. A reader might bypass this section on first reading, unless a particular sensor is very important to an application of current interest. Sensor technology is advancing rapidly; we should expect not only new sensors in the future, but also better performance of current devices.

* Microdensitometer

Slides or film can be scanned by passing a single beam of light **through** the material: a single sensor on the opposite side from the light records the optical density of the material at location **[r, c]**. The material is moved very precisely by mechanical stages until an entire

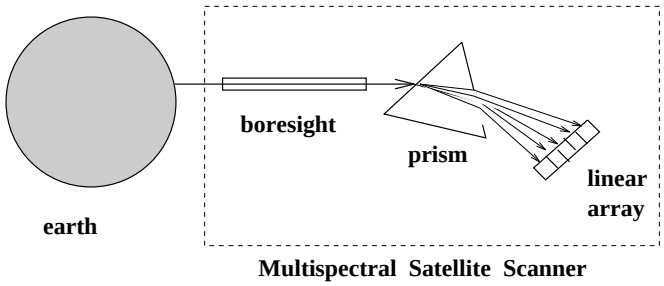

Figure 2.15: Boresighted multispectral scanner aboard a satellite. Radiation from a single surface element is refracted into separate components based on wavelength.

rectangular area is scanned. Having a single sensor gives one advantage over the CCD array, there should be less variation in intensity values due to manufacturing differences. Another advantage is that many more rows and columns can be obtained. Such an instrument is slow however and cannot be used in automation environments.

The reader might find some interest in the following history of a related scanning technique. In the 1970's in the lab of Azriel Rosenfeld, many pictures were input for computer processing in the following manner. Black and white pictures were taken and wrapped around a steel cylinder. Usually 9 x 9 inch pictures or collages were scanned at once. The cylinder was placed in a standard lathe which spun all spots of the picture area in front of a small LED and sensor that measured light reflecting off each spot. Each revolution of the cylinder produced a row of 3600 pixels that were stored as a block on magnetic tape whose recording speed was in sync with the lathe! The final tape file had 3600 x 3600 pixels, usually containing many experimental data sets that were then separated by software.

*** Color and Multispectral Images**

Because the human eye has separate receptors for sensing light in separate wavelength bands, it may be called a *multispectral* sensor. Some color CCD cameras are built by placing a thin refracting film just in front of the CCD array. The refracting film disperses a single beam of white light into 4 beams falling on 4 neighboring cells of the CCD array. The digital image it produces can be thought of as a set of four interleaved color images, one for each of the 4 differently refracted wavelength components. The gain in spectral information is traded for a loss in spatial resolution. In a different design, a color wheel is synchronously rotated in the optical pathway so that during one period of time only red light is passed; then blue, then green. (A color wheel is just a disk of transparent film with equal size sectors of each color.) The CCD array is read 3 times during one rotation of the color wheel to obtain 3 separate images. In this design, sensing speed is traded for color sensitivity; a point on a rapidly moving object may actually image to different pixels on the image plane during acquisition of the 3 separate images.

Some satellites use the concept of *sensing through a straw or boresight*: each spot of the earth is viewed through a *boresight* so that all radiation from that spot is collected at the same instant of time, while radiation from other spots is masked. See Figure 2.15. The beam of radiation is passed through a prism which disperses the different wavelengths onto

a linear CCD array which then simultaneously samples and digitizes the intensity in the several bands used. (Recall that light of shorter wavelength is bent more by the prism than light of longer wavelength.) Figure 2.15 shows a spectrum of 5 different bands resulting in a pixel that is a vector $[\ b_1, b_2, b_3, b_4, b_5\ ]$ of 5 intensity values. A 2D image is created by moving the boresight or using a scanning mirror to get columns of a given row. The motion of the satellite in its orbit around the earth yields the different rows of the image. As you might expect, the resulting image suffers from motion distortion – the set of all scanned spots form a trapezoidal region on the earth whose form can be obtained from the "rectangular" digital image file using the warping methods of Chapter 11. By having a **spectrum of intensity values** rather than just a single intensity for a single spot of earth, it is often possible to classify the ground type as water or forest or asphalt, etc.

*** X-ray**

X-ray devices transmit X-ray radiation through material, often parts of the human body, but also welded pipes and jars of applesauce. Sensors record transmitted energy at image points on the far side of the emitter in much the same manner as with the microdensitometer. Low energy at one sensed image point indicates an accumulation of material density along the entire ray projected from the emitter. It is easy to imagine a 2D X-ray film being exposed to X-rays passing through a body. 3D sensing can be accomplished using a CT scanner ("cat" scanner), which mathematically constructs a 3D volume of density values from data collected by projecting X-rays along many different rays through the body. At the right in Figure 2.16 is a 2D computer graphic rendering of high density 3D voxels from a CT scan of a dog: these voxels are rendered as if they were nontransparent reflecting surfaces seen from a particular viewpoint. A diagnostician can examine the sensed bone structure from any viewpoint.

Exercise 11

Think about some of your own dental X-rays: what was bright and what was dark – the dense tooth or the softer cavity? Why?

*** Magnetic Resonance Imaging (MRI)**

Magnetic resonance imaging (MRI) produces 3D images of materials, usually parts of the human body. The data produced is a 3D array **I[s, r, c]**, where **s** indicates a *slice* through the body and **r** and **c** are as before. Each small volume element or *voxel* represents a sample perhaps 2 mm in diameter and the intensity measured there is related to the chemistry of the material. Magnetic resonance angiography (MRA) produces intensities related to the speed (of blood flow) of material at the voxel. Such scanners can cost a million dollars and a single scanning can cost a thousand dollars, but their value is well established for diagnosis. MRI scanning can detect internal defects in fruits and vegetables and may be used for such in the future if the cost of the device drops. Figure 2.16(left) shows a digital image extracted from 3D MRA data. This *maximum intensity projection*, or $MIP[r, c]$, could be produced by choosing the brightest voxel $I[s, r, c]$ over all slices s. A computer algorithm can actually generate a MIP image by projecting in any view direction. Diagnosis is typically done using a wall full of such printed 2D images, but true 3D displays are now available and radiologists are learning to use them.

Figure 2.16: (Left) A maximum intensity projection (MIP) made by projecting the brightest pixels from a set of MRA slices from a human head (provided by MSU Radiology); (right) a computer generated image displaying high density voxels of a set of CT scans as a set of illuminated surface elements. Data courtesy of Theresa Bernardo.

*** Range Scanners and Range Images**

Devices are available that sense depth or range to a 3D surface element, rather than just the intensity of radiation received from it. Samples of the surface shape of objects are directly available in a range image, whereas in an intensity image, surface shape can only be derived from difficult and error prone analysis. A LIDAR device, shown in Figure 2.17, transmits an amplitude modulated laser beam to a spot of the 3D surface and receives the reflected signal back. By comparing the change of phase (delay) between the transmitted and received signal, the LIDAR can measure the distance in terms of the period of the modulation of the laser beam. This works only for distances covered by one period because of the ambiguity – a spot at distance of $d + n\lambda/2$ produces the same response as a spot at distance d, where λ is the period of modulation. Moreover, by comparing the received intensity to the transmitted intensity, the LIDAR also estimates the reflectivity of the surface spot for this wavelength of laser light. Thus, the LIDAR produces two registered images – a range image and an intensity image. The LIDAR is slower than a CCD camera because of the *dwell time* needed to compute the phase change for each spot: it is also much more expensive because of the mechanical parts needed to steer the laser beam. This expense has been justified in mining robots and in robots that explore other bodies in our solar system.

A variation of the 5000 year-old surveying method of triangulation can be used to obtain 3D surface measurements as shown in Figure 2.18. A plane of light is projected onto the surface of the object and the bright line it produces is observed by a camera. Each bright image point $[\ x_c, y_c\ ]$ is the image of some corresponding illuminated 3D point $[\ x_w, y_w, z_w\ ]$. So, the sensing device "knows" the plane of light and the ray of light from the camera center through the image point out into 3D space. From intuitive geometry, we know that the imaging ray will pierce the plane of light in a unique point. The coordinates x_w, y_w, z_w

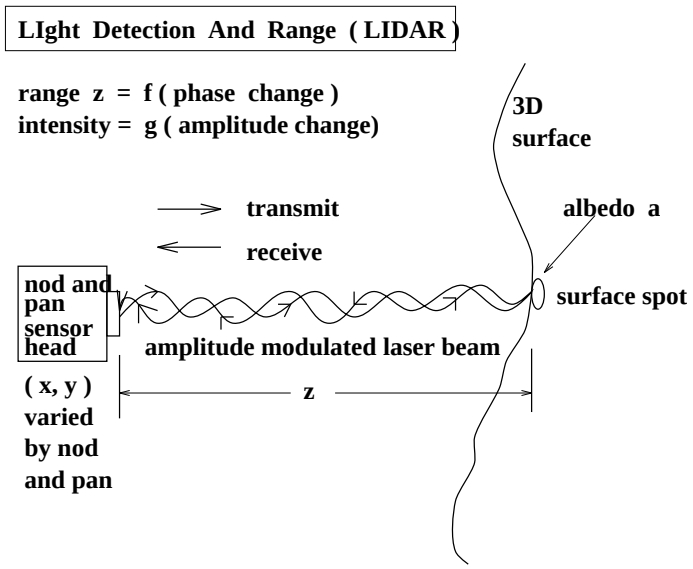

Figure 2.17: A LIDAR sensor can produce pixels containing both range and intensity.

can be determined by analytical geometry: we have one equation in those 3 unknowns from the light sheet and 2 equations in those 3 unknowns from the imaging ray; solving these 3 simultaneous linear equations yields the location of the 3D surface point. In Chapter 13, *calibration* methods are given that enable us to derive the necessary equations from several measurements made on the workbench.

The above argument is even simpler if a single beam of light is projected rather than an entire plane of light. Many variations exist and a sensor is usually chosen according to the particular application. To scan an entire scene, a light sheet or beam must be swept across the scene. Scanning mirrors can be used to do this, or objects can be translated past the sheet over time using a conveyor system. Many creative designs can be found in the literature. Machines using multiple light sheets are used to do automobile wheel alignment and to check the fit of car doors during manufacture. When looking at specific objects in very specific poses, image analysis may just need to verify if a particular image stripe is close enough to some ideal image position. The stream of observations from the sensor is used to adjust online manufacturing operations for quality control and for reporting offline.

2.10 References

More specific information about the design of imaging devices can be found in the text by Schalkoff (1989). Tutorials and technical specifications of charge-coupled devices are readily found on the web using a search engine: one example is a tutorial provided by the University of Wisconsin at www.mrsec.wisc.edu/edetc/ccd.html. One of several early articles on the early development of color CCD cameras is by Dillon *et al* (1978). Discussion and modeling of many optical phenomena can be found in the book by Hecht and Zajac (1976).

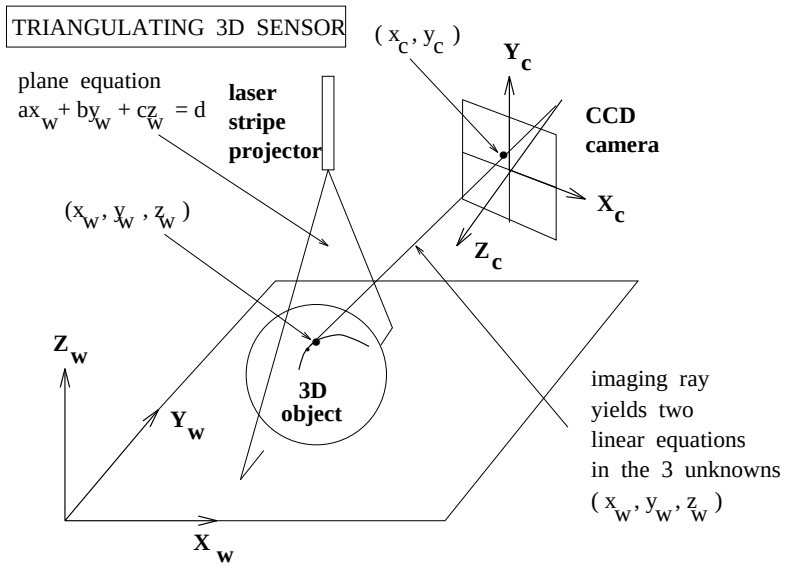

Figure 2.18: A light striping sensor produces 3D points by triangulation.

Many fundamental observations leading to computer vision techniques can be found in the book by pyschologist J.J. Gibson (1950). Properties of animal vision systems and the human visual system from the perspective of an engineer are given in the text by Levine (1985). The book by Nalwa (1993) begins with a discussion of the capabilities and faults of the human visual and gives a good intuitive description of imaging and the perspective transformation. Margaret Livingston (1988) gives a popular treatment of human perception with an orientation to art appreciation. Many known mathematical properties of the perspective transformation are contained in Haralick and Shapiro, Volume II (1992). Practical details and an integrating overview on image file formats is provided in the encyclopedia by Murray and VanRyper (1994); a CD of common software utilities collected from several different sources is included.

1. P. Dillon, D. Lewis and F. Kaspar,*Color Imaging System using a Single CCD Area Array*, IEEE Transactions on Electron Devices, Vol. ED-25, No. 2 (Feb 1978)102-107.

2. J.J. Gibson (1950), **The Perception of the Visual World**, Houghton-Mifflin, Boston.

3. E. Hecht and A. Lajac (1974), **Optics**, Addison-Wesley.

4. R. Haralick and L. Shapiro (1992), **Computer and Robot Vision, Volumes I and II**, Addison-Wesley.

5. M.D. Levine (1985), **Vision in Man and Machine**, McGraw-Hill.

6. M. Livingstone, (1988) *Art, Illusion and the Visual system*, Scientific American, Jan. 1988, 78-85.

7. J. Murray and W. VanRyper (1994), **Encyclopedia of Graphics File Formats**, O'Reilly and Associates, Inc., 103 Morris St., Suite A, Sebastopol, CA 95472.

8. V. Nalwa (1993), **A Guided Tour of Computer Vision**, Addison-Wesley.

9. R.J. Schalkoff (1989), **Digital Image Processing and Computer Vision**, John Wiley and Sons.